基于 DEA 超效率验证的项目同行评议结果有效性研究

万　昊[1]，张福俊[2]，吕千千[3]
（1.山东科技大学 图书馆，青岛　266590；
2.山东科技大学 发展规划处，青岛　266590；3.中国科学院文献情报中心，北京　100190）

摘　要：[目的 / 意义]本文以项目评审为例探索专家知识为基础的同行评议判断的有效性。[方法 / 过程]本文采用数据包络分析测度项目的研发效率，评价模型为经超效率修正的 CCR 模型，并将该效率指标作为同行评议有效性的验证标准。数据源为国家林业局实施完毕的 126 项“948”项目，涵盖 13 项投入与产出评价指标。[结果 / 结论]实验发现：项目同行评议结果与实际研发效率（DEA 超效率）之间呈现适度的正相关性，Spearman 相关系数 ρ=0.250（*Sig.*<0.01）。主观评审犯Ⅰ类错误与Ⅱ类错误的比重分别为 17.5%与 18.2%，有效同行评议决策比重占 64.3%。证实同行评议的确存在一定的系统误差但不可避免。

关键词：同行评议；项目评价；有效性；超效率 DEA

中图分类号：G301；G350　　**文献标识码：**A　　**文章编号：**1002-1248（2022）02-0088-14

引用本文：万昊，张福俊，吕千千．基于 DEA 超效率验证的项目同行评议结果有效性研究[J]．农业图书情报学报，2022，34（2）：88-101．

1 引　言

现代同行评审制度起源于 18 世纪，当时英国皇家协会（科学共同体）在全球刚成立，传统的书信交流模式被期刊交流模式取代，同时以共同体内部成员同行评议的方式对学术质量监督开始确立[1]。目前，同行评议机制在学界广受赞誉，其应用领域扩展到：①科学家智力劳动成果（论文手稿）发表的评审；②科学家岗位雇佣和职业晋升（由副教授晋升为教授）；③科学家颁奖（如诺贝尔奖、菲尔兹奖、图灵奖）评审；④项目经费资助的审批；⑤学科发展态势评估及研发评价实践（Research Assessment Exercise，RAE）；⑥F1000 等专家的论文推荐打分[2]。发展至今，同行评议为基础的共同体内部成员间的学术质量监督机制正逐渐发展成为一项系统性工程，来维护科学交流的高水平和高质量。

长久以来，科学共同体成员坚持科学内部有其内在的客观标准，科学家坚持这一标准开展同行评议。该无形标准实际上为科学社会内部成员一致认可并且

收稿日期：2021-07-08

基金项目：山东科技大学人才引进科研启动基金项目“同行评议与计量评价在科研评价中应用比较研究”（2017RCJJ094）；国家林业局 948 计划项目“引进林业产业技术评估、集成与创新”（2012-4-25-2）

作者简介：万昊，男，博士，馆员，研究方向为科技评价与创新性评估。张福俊，男，教授，硕士生导师，山东科技大学发展规划处，副处长，研究方向为数据科学与知识可视化技术。吕千千，女，博士研究生，研究方向为科技评估

普遍遵守的自律法则，规范科学家群体的研究行为，使得同行评议的结果收敛并达成一致性共识。这也是莫顿（MERTON）科学社会学中指出的科学内在的民主有序性的根本原因[3]。

但是，同行评议在评价实践中的广泛使用同时，其可能的不完备一面正受到越来越多的关注（主观性偏差）。荷兰国家研究委员会经费资助、澳大利亚研究委员会经费资助、美国国家科学院报告对同行评议有偏源都有报道[2]。为了避免主观有偏对评审结果产生的影响，PLosOne、F1000 和 Publons 平台相继推出开放的同行评议，避免双盲评审可能的偏差[4,5]。

目前来看，任何量度工具都有自身的局限。同行评议由于专家主观性，会导致评审结果的嵌入式有偏[6]。研究人员不禁要问：同行评议遴选结果在多大程度上是有效的或有价值的？同行评议在多大比例上实现其历史使命来遴选出最有前景和应用价值最好的项目来实现资源高效配置？同行评议在多大比例上能够保证稀缺的经费资源或版面资源分配给高质量的研究？本文以项目资助评审为实证探究。

2 文献综述

2.1 国内外相关研究

WESSELY 和 WOOD[7]指出同行评议方法它不仅是一种学术性的探索，更是一种政策性的诉求，致力于通过客观、公正的评价来维护科学系统的健康运转并提升科学研究的质量。而对科技项目资助评审领域的调研结果给出 3 类“异质性”答案。

2.1.1 适度的正相关

MAVIS 和 KATZ[8]统计美国畸形儿基金会、BORNMANN[9]统计了德国勃林格·殷格翰基金会（BIF）、BORNMANN 等[10]调研欧洲分子生物机构（EMBO）、NEUFELD 和 VON INS[11]调研德国研究基金会（DFG）、LI 和 AGHA[12]调研美国国立卫生研究院（NIH），结果发现：资助与未资助的科研项目的论文产出数量、引用量和篇均被引量，前者分别是后者的1.8~3.1 倍。受资助项目的文献计量表现显然更佳，进而佐证了评审委员会的专家评审结果与文献计量结果之间存在显著的正相关性。

2.1.2 不相关

MELIN 和 DANELL[13]对瑞典战略研究基金会（SFSR）、HORNBOSTEL 等[14]汇总德国研究基金会（DFG）埃米·诺特计划（ENP）年轻科学家的奖学金资助申请对比发现：成功和失败的申请者在经费资助审批前 n 年与后 n 年的年均出版量、篇均引文量和发文期刊的影响因子等文献计量指标差不多。

2.1.3 正相关与负相关的混合结论

VAN DEN BESSELAAR 和 LEYDESDORFF[15]统计了荷兰国家科学研究委员会（NRC）中经济和社会行为科学两个学部、BORNMANN 等[16]对荷兰经济和社会科学研究委员会（MaGW）项目资助的申请者调研，结果发现：某些被拒绝的研究人员也会比接受基金资助的研究人员有更好的科研绩效表现（出版和引用量），学术声誉低的申请者可能提出更高价值的科学研究，而学术声誉高的申请者也可能从事并不具有创新价值的研究，二者量度结果不一致。

2.2 本文切入点

得出异质性的结果，受很多因素的影响。首先，科研产出的成果数量与学术影响力可能受到科学家年龄的影响，对于一些青年研究资助项目或博士后项目资助，受资助成功者与失败者的科研表现差距并不明显。其次，没有共通一致认可研究标准、测度指标（时间窗）、变化指标、实验步骤、流程设计，不便于不同研究者结果对比。

目前，同行评议项目资助评审的有效性验证标准普遍采用引用为基础的测度指标（如被引量、期刊影响因子 JIF、Citescore 值、SNIP 指数、SJR 指数、H 指数、皇冠指数）作为同行评议决策（学术水平）的验证标准。引用测度仅仅适合于基础研究以及哲学人文社科类的研究，而对于应用性比较强的项目进行量度评价就不再适用了。因为基础研究的成果仅仅是论文，应用研究项目的产出成果还涉及专利与标准等。英国

1975—1992 年在心血管和中风疾病研究发现，平均每 ₤1（US $1.43）投入将会带来在经济和社会健康领域 ₤1.39 的回报[17]。2010 年，美国国立卫生研究院（NIH）现任院长 COLLINS 在健康宣传报告中指出：美国国立卫生研究院支出的每$1 都会在 12 个月内转化成为$ 2.21 的额外经济产出[18]。

考虑基础研究投入的多样性回报，为了在更广的维度上考虑项目的多样性产出，本文选择数据包络分析（DEA）方法，许多研究将 DEA 方法应用到科研投入 / 产出的计量分析中，尤其科技管理与科技政策领域，它有诸多优点：① 对数据源没有要求，不需要服从一定的假设分布；② 无需假定生产函数来估计生产前沿面；③ 可解决多投入 / 多产出的效率计算；④ 非参数模型，不需要考虑投入 / 产出指标的权重；⑤ 不需要考虑指标量纲。本文通过更加先进的 DEA 模型（指标）来测度项目的科研生产率（表征学术价值），并将其作为同行评议决策有效性的验证标准，而非引用量度（表征学术影响力），以求通过更加精准模型来印证定性同行评议价值判断与定量评价模型的量度结果是否相一致？同行评议决策过程中犯 I 类与 II 类错误及其比例？

3 理论分析与实验设计

3.1 原理：DEA 模型

数据包络分析（Data Envelopment Analysis，DEA）是运筹学和经济学中用于估计生产前沿面的非参数方法。它在实证研究中用于测量被评价单元或决策单元（Decision Making Unit，DMU）的生产效率，DMU 指将投入转化为产出的创新主体。

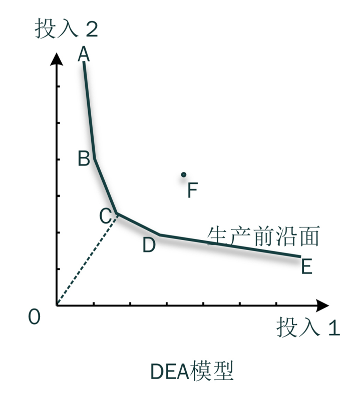
3.1.1 传统 DEA 模型

数据包络分析法作为非参数法，借助线性规划思想，应用实证方式估计生产技术的前沿面，并测量“决策单元的效率”，该模型被命名为 CCR 模型[19]。本文评价模型也是选择投入角度（Input-oriented）的 CCR 模型，即将线性规划在产出不变的条件下既定的生产要素生产一定量产品所需要花费的最小成本占实际成本的比率，如果该比率为 1，则称 DMU 技术有效。

3.1.2 超效率(SE)-DEA 模型

DEA 超效率模型作为 DEA 方法的一种改进。传统 DEA 模型依据决策单元是否位于“生产前沿面”上判断决策单元是否有效，但却无法判断有效决策单元的优劣。为了弥补这一缺陷，ANDERSEN 和 PETERSEN 提出了超效率（Super Efficiency，SE）模型，超效率 SE-DEA 模型能够对有效前沿面上的决策单元进行再排序。其基本思想是：在评估决策单元（DMU）时，将该决策单元本身排除在单元集合之外。为了更直观地说明以图 1 来阐述超效率模型的原理。

传统 CCR 模型计算的效率值是将 DMU 与包络面上的有效 DMU 相比，因此可以区分有效与无效 DMU，而对于有效的 DMU 进一步区分需要超效率 SE-DEA 模型。DEA 模型中，C 点由于位于生产前沿面 ABCDE 上，C 点技术有效，效率值 θ_C=1；超效率 SE-DEA 模型中，计算 C 点技术效率要将该点排除在



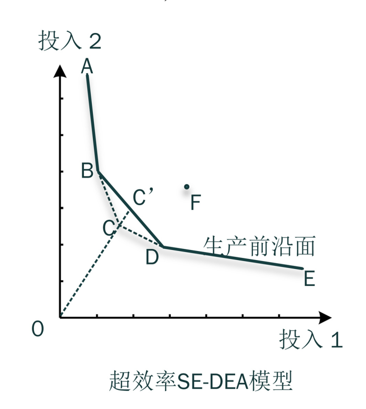


图 1　传统 DEA 模型与超效率 SE-DEA 模型对比（投入角度）

Fig.1 Comparison between traditional DEA model and super-efficiency SE-DEA model (input-oriented)

集合外，生产前沿面变成 ABDE 上，C 点技术效率 θ_C^{super}=OC'/OC>1。

超效率 DEA 评价模型与传统 DEA 模型的数学形式基本相似，对于线性规划 D'_{CCR} 有如下定义：设有 n 个决策单元（DMU），每个 DMU 的有 m 种类型输入，有 s 种类型输出。对于 DMU_j（$j\in[1,\cdots,n]$），有：

X_j 和 Y_j 分别表示第 j 个决策单元的输入向量和输出向量；

X_{ij}=第 j 个决策单元的第 i 种输入的投入量，$x_{ij}>0$（$1<i<m$）；

Y_{rj}=第 j 个决策单元的第 r 种输出的产出量，$y_{rj}>0$（$1<r<s$）；

$$D'_{CCR}\begin{cases} min\ \ [\theta_{super}-\varepsilon(\sum_{i=1}^{m}s_i^-+\sum_{r=1}^{s}s_r^+)] \\ s.t.\ \sum_{j=1}^{n}\lambda_jX_{ij}+s_i^-=\theta_{super}X_i,\quad i=1,2,\ldots,m \\ \sum_{j=1}^{n}\lambda_jY_{rj}-s_r^+=Y_r,\quad r=1,2,\ldots,s \\ \lambda_j\geqslant 0,\quad j=1,2,\cdots,n \\ s_i^-\geqslant 0,\quad s_r^+\geqslant 0 \end{cases}$$

在线性规划模型中，θ_{super} 为决策单元 DMU 的超效率值；s_i^-与 s_r^+为松弛变量，分别表示投入冗余与产出不足，ε 为非阿基米德无穷小量。模型中，当 $\theta_{super}>1$ 时，决策单元的科研效率为 DEA 有效，达到了最优解；当 $\theta_{super}=1$，决策单元为（弱)DEA 有效；当 $\theta_{super}<1$ 时，决策单元不是 DEA 有效，经济活动不是技术效率最佳。

3.2　评价指标设计

而针对科研项目评价后验性有效性分析的指标体系设计，本文借鉴欧盟委员会[20]对科学研究项目（尤其是应用性研究）评价，突破传统的仅基于基础性学术论文产出的量度模式，进一步拓展到整个研发活动的全产业链。

详细地说，一方面，项目投入评价，其指标涵盖研发经费 I1、研发人员 I2、研发设备 I3、研发周期 I4、技术培训 I5 共 5 项指标。另一方面，项目产出评价。根据 OECD 对项目类型的划分（基础研究、应用研究、试验开发类研究），科研产出成果不仅涵盖：①哲学、人文社科类的基础研究普遍的学术著作 O1、期刊论文 O2、研究报告 O3 等学术交流出版物；也包括②自然、理工类学科由于知识的转移所带来的技术性产出，由专利 O4、标准 O5、计算机软件、加工工艺等其它保密性成果 O6 表征；③甚至随着技术逐渐成熟，从实验室阶段过渡到中试线与实验基地 O7 流程化生产阶段；④随着技术成熟度的进一步提高，技术成果进一步向下游转移、转化，由实验室样品到产业化生产的商品，所带来的商业利润等直接与间接回报 O8；⑤同时，技术产出在国内外的先进水平由 O9 表征；⑥而且，该知识创造、技术转移、科技成果转化过程密切了大学与企业之间的交流与合作，因此在产出评价指标之中也必须考虑科研合作频次 O10[21]；⑦人才在联系“大学与企业”技术转移转化过程中的关键性纽带作用，根据 SALTER 和 MARTIN 的研究：拥有科学理论武装的优秀大学毕业生和技能熟练的工程师进入产业部门工作来解决复杂的科学与技术问题，是基础研究转化成为经济获益的主要途径[22]。换句话说，培养训练有素的问题解决者（Trained Problem-Solvers）是联系大学与企业之间的桥梁，培养有资格的科学家和工程师作为知识储备的可靠资源池，本文用人才产出 O11 该指标表征（表 1）。

3.3　案例数据：“948 计划”

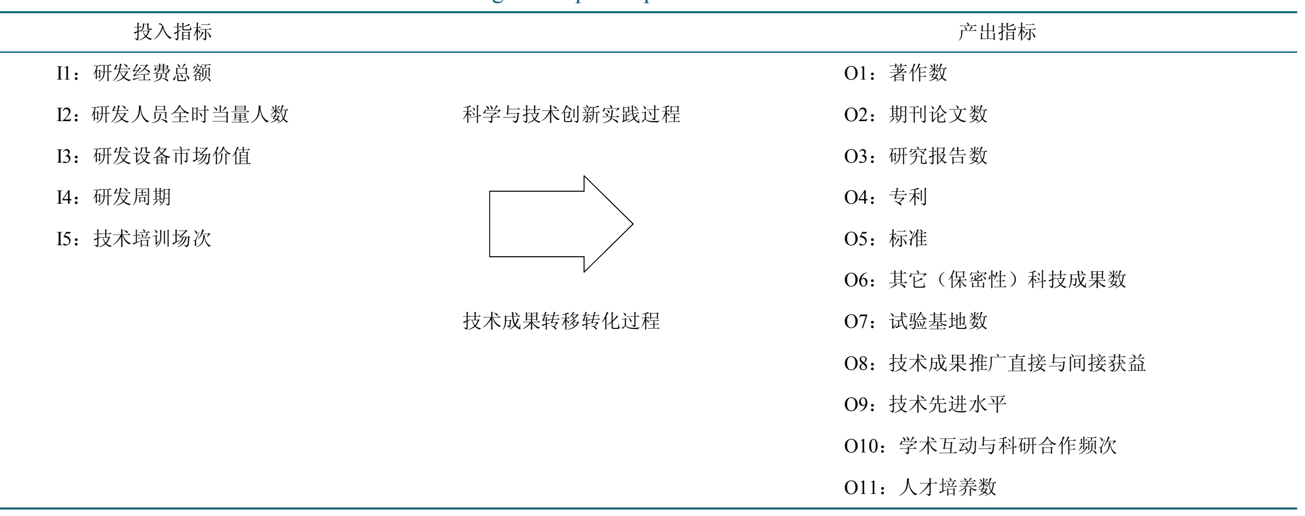
“948 计划”是中国唯一对国际先进的农、林、水先进技术引进的专项计划。经国务院批准，该计划自 1996 年 5 月开始实施，以项目合同制的形式承包给大学和科研机构。948 计划作为技术引进类项目，以应用为导向，其成果既包括论文学术产出，还包括专利技术产出，甚至技术转移、转化与推广应用获得的经济利益，涉及整个研发产业价值链所有环节可能的成果产出类型，本研究将其作为项目的数据源（表 2）。

3.4　实验过程

本文实验过程技术路线，如图 2 所示。

表 1　DEA 模型的“投入 - 产出”指标设计

Table 1 Design of "input-output" indicators of DEA model



投入指标		产出指标
I1：研发经费总额		O1：著作数
I2：研发人员全时当量人数	科学与技术创新实践过程	O2：期刊论文数
I3：研发设备市场价值		O3：研究报告数
I4：研发周期		O4：专利
I5：技术培训场次		O5：标准
		O6：其它（保密性）科技成果数
	技术成果转移转化过程	O7：试验基地数
		O8：技术成果推广直接与间接获益
		O9：技术先进水平
		O10：学术互动与科研合作频次
		O11：人才培养数

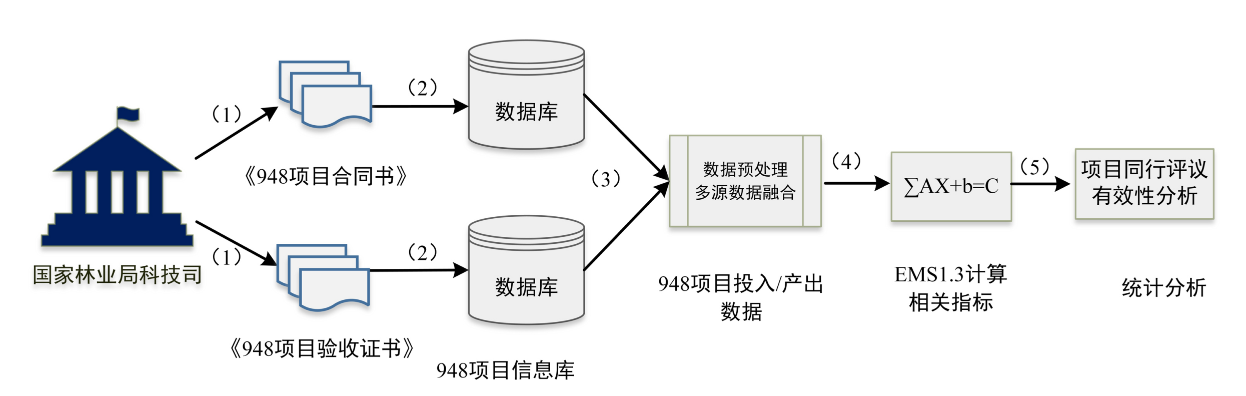


图 2　本文实证研究的实验过程

Fig.2 Experimental research process in this paper

步骤（1）：数据获取。由于“948 计划”项目的申请、资助、评审与验收文档资料属于不公开的灰色文献，国家行政部门内部资料。本文数据选取国家林业局承担实施的“948 计划”项目，借助于承担项目便利提取 1996 年以来存放于国家林业局科技司的纸质版 948 项目《合同书》与《验收证书》。

步骤（2）：数字化加工处理。将纸质版的 948 项目《合同书》和《验收证书》文档进行数字化处理，获取电子版 PDF 原文。进一步开展二次化深加工，从项目的实施内容、考核指标和成果产出等信息项中提取相应的元数据（论文、专利、标准、认定成果、示范点等），采用统一的资源描述框架规范将非结构化的文本信息转换成结构化的项目信息并录入到“948 林业项目信息库”中，进而发布在中国林业信息网（http://www.lknet.ac.cn）林业成果栏下以方便用户对项目信息检索。

步骤（3）：实验数据的提取与融合。由于 2000 年以前国家林业局实施的 948 计划项目规模较少，在 2001—2011 年间批准实施 948 计划项目共 764 项，涵盖生态建设类项目 402 项与林业产业技术类项目 362 项。考虑到项目的实施周期，本文仅提取 2001—2005 年（也就是在“十五”期间）批准实施的 126 项林业产业技术类项目（简称“林产项目”），在此期间的项目全都实施完毕并通过了结题验收。输入数据的预处理。按照本文确立的评价指标体系，将提取的 126 项 948 项目的投入与产出信息。这其中涉及到合同信息与验收信息的异构数据融合。

步骤（4）：效率值计算。本文采用 EMS 1.3 经济

表 2 “十五”期间批准实施的 948 林产技术项目的评价指标信息

Table 2 Evaluation indicators of "948 project" forest products technologies approved during the "Tenth Five-Year Plan" period

序号	项目编号	项目名称	承担单位	R&D经费/万元{I1}	研发团队人数/人{I2}	研发设备市值/万元{I3}	研发持续时间/年{I4}	……	出版著作/部{O1}	期刊论文/篇{O2}	研究报告/个{O3}	专利/个{O4}	标准/个{O5}	……	学术互动与科研合作/项{O10}	人才培养/人{O11}
1	2001-01(1)	木塑复合材料的注塑成型技术引进	中国林科院木材工业研究所	99.5	5	67.5	3	……	0	6	0	2	0	……	3	1
		……	……											……		
16	2002-19	加拿大旋切单板柔化技术及关键设备引进	南京林业大学	70	6	0	3	……	0	2	0	2	1	……	1	0
17	2002-27(1)	梨果产后商品化处理及包装保鲜技术引进	河北省林业科学研究院	50	16	26.5	5	……	0	5	0	0	0	……	2	0
18	2002-34	病虫害防治低量风送高射程喷雾技术引进	南京林业大学	65	18	26.7	5	……	0	6	0	15	1	……	2	4
19	2002-36	气助式静电喷雾机技术引进	国家林业局森林病虫害防治总站	60	6	1	3	……	0	4	1	1	0	……	3	0
20	2002-39(1)	便携式木材智能无损探测仪的引进	东北林业大学	50	19	35.1	3	……	0	13	0	1	0	……	2	8
21	2002-39(2)	原木 X 射线无损检测系统的引进	东北林业大学	40	9	0	3	……	1	19	1	1	0	……	2	0
		……	……					……						……		
114	2005-4-68	软木热-化学膨胀技术引进	西北农林科技大学	50	12	10.4	3	……	0	7	0	3	0	……	2	4
115	2005-4-69	竹醋液深度开发关键技术引进	浙江林学院	90	7	7.5	3	……	2	4	0	2	4	……	5	0
116	2005-4-70	基于纤维素及其衍生物的高分子乳液制备技术引进	中国林科院林产化学工业研究所	40	10	0	3	……	0	2	0	4	0	……	2	1
		……	……					……						……		
125	2005-4-C04	濒危野生动物基因鉴定试剂盒研制技术创新	国家濒危野生动植物种质基因保护中心（浙江大学）	140	7	2.2	2	……	0	7	0	0	0	……	0	0
126	2005-4-C06	紫胶漂白胶水果保鲜剂研制及应用技术创新与示范	中国林科院资源昆虫研究所	100	11	32.3	3	……	0	11	0	7	0	……	0	0

* 注：鉴于版面有限，这里仅提供局部 948 项目信息

计量软件，按照软件的要求进行规范化的整理，模型选择投入导向的 CCR 模型，计算决策单元的 DEA 超效率（SE）值。

步骤（5）：数据分析与绘图。借助 SPSS 20.0 对项目的同行评议打分与 DEA 超效率（SE）开展正态分布检验与相关性分析，以及两指标与典型投入 / 产出指标的相关性分析。绘制以上两指标的散点分布图并计算 Ⅰ 类误差（纳伪）和 II 类误差（弃真）的比重。

4 研究结果

4.1 超效率(SE)–DEA 模型测度结果分析

本文将计算的项目 DEA 超效率值按照降序排列（表 3）。结果发现：①研发效率高 Top 10 项目，主要分布在“木材加工”“林产化工”领域，这些项目实施涉及研发全产业链，不仅有论文还有技术产出，甚至技术成熟到中试阶段，并有一定的市场获益。如林科院前院长江泽慧、林科院木材所前所长费本华主持的项目，其项目依托单位是科研院所，有丰富的师资、教育、国际合作等研发创新的条件与基础，所以项目实施有很高的研发效率。②而依托单位为国家林业局由于没有事业单位的条件基础，又缺乏团队与研究生支撑，脱离了教学与科研一线，所以项目的研发效率值偏低，如周建仁处长、王琦处长从林科院调到林业局工作属于该情况。③还有一些项目更贴近上游基础性实验，新颖性高但实用性低、成本高但技术成熟度低，这导致市场推广前景不大，同时没有论文与专利成果，最终导致项目的研发效率值较低，如卢孟柱教授、张宗和教授承担项目皆属于该情况。

4.2 同行评议决策的有效性

在同行评议决策有效性的界定上，我们借鉴经典论述[14]：以同行评审打分与研发效率（DEA 超效率）的中位数为分割点，将整个二维联合分布划分为 4 象限。其中，同行评审打分高于平均水平而其科研效率低于平均水平的项目，为同行评审的 I 类误差；而同行评审打分低于平均水平而其科研效率高于平均水平的项目，为同行评审的 II 类误差（表 4）。

我们将计算的项目同行评审值与 DEA 超效率值汇总到 XY 直角坐标系中，并借助 SPSS 绘制 X 轴与 Y 轴上的频次分布图。以 DEA 超效率的中位数点所在轴线为 X 轴，以同行评议得分的中位数点所在轴线为 Y 轴，将整个 XY 直角坐标系划分为 4 个象限，其中左上象限点为被高估的 I 类误差（“纳伪”）分布点，右下象限点为被低估的 II 类误差（“弃真”）分布点（图 3）。

本研究发现：同行评议正确决策的比重 64.3%，Ⅰ类误差的比重为 17.5%，Ⅱ类误差的比重为 18.2%（表 5）。

而在已有的项目经费资助评审的有效性研究中，如 BORNMANN 对德国勃林格·殷格翰基金会的博士后奖学金资助[9]以及对欧洲分子生物学组织的奖学金和青年研究者计划项目资助[10]，结果皆发现：II 型误差的比重远大于 I 型误差。本文 948 资助计划的结果不同：I 类和 II 类误差的比重大致相当。该差异原因：欧洲、美国的项目资助评审更为严格，有盲评、会评等流程，而且评审专家组成员也有严格的遴选机制、打分有严格的标准，评审质量也更高，所以专家打分更倾向于犯 II 类错误（弃真）而非 I 类错误（纳伪），即为实现高水平项目的有效遴选“宁可错杀一千也不放过一个”。而国内项目基金会的运作环境较为宽松。启示：更应像西方一样加强“过程质量监管”，严格层层把关，扭转中国当前“重立项轻验收”的局面。

正态分布检验，在消除样本量和异常值因素对统计分析方法选择造成影响与干扰基础上，无论 Kolmogorov-Smimov 检验和 Shapiro-Wilk 检验两种方法的结果一致：同行评议得分（Scores）的伴随概率分别为 $P_{K\text{-}S}$=0.200 与 $P_{S\text{-}W}$=0.067，皆大于显著性水平 0.05，接受零假设，服从正态分布；DEA 超效率（SE）的伴随概率分别为 $P_{K\text{-}S}$=0.003 与 $P_{S\text{-}W}$=0.000，皆小于显著性水平 0.05，拒绝零假设，不服从正态分布（表 6）。

4.3 统计结果的相关性检验

根据正态检验结果（表 7），DEA 超效率（SE）的

表 3 “十五”期间实施完毕的 126 项 948 林产项目的量化（含 DEA 超效率）结果

Table 3 Quantification results (including DEA super-efficiency) of 126 "948 forestry projects" completed during the "Tenth Five-Year Plan" period

序号	项目编号	项目名称	承担单位	专业分类	项目负责人	引进国家地区	起止时间	验收证书编号	验收评分	研发经费/万元{I1}	期刊论文/篇{O2}	专利/个{O4}	DEA-超效率
76	2004-4-55	木材机械应力分级技术引进	中国林科院木材工业研究所	木材加工	任海青；费本华	加拿大；美国	2004—2007 年	林科验 2008 年 118 号	83.2	60.0	28	0	15.000
110	2005-4-64	油橄榄叶片提取物及渣油萃取技术引进	甘肃省林业科学研究院	林产化工	于洪波	美国；希腊	2005—2008 年	林科验 2008 年 024 号	81.9	70.2	5	1	5.000
90	2005-4-08	干旱区机械化集雨植被恢复关键技术引进	国家林业局调查规划设计院	林业机械	王军厚；赵有贤	意大利	2005—2008 年	林科验 2008 年 050 号	94.6	40.0	1	0	4.500
13	2001-34	麦（稻）秸纤维制备技术引进	南京林业大学	生物基材料	周定国	奥地利；德国	2001—2004 年	林科验 2006 年 045 号	88.4	80.0	80	4	3.509
80	2004-4-63	低质材高效低污染制高档纸浆生产技术引进	中国林科院林产化学工业研究所	林产化工	房桂干	奥地利	2004—2007 年	林科验 2008 年 123 号	88.7	60.0	18	0	2.192
38	2003-4-28	木结构房屋结构材料应用关键技术引进	中国林科院木材工业研究所	木材加工	费本华；吕建雄	加拿大；日本	2003—2008 年	林科验 2008 年 014 号	87.3	333.0	26	8	2.015
96	2005-4-22	城市森林监测、评价与标准化数据收集技术引进	中国林科院林业科技信息研究所	林业机械	李智勇	丹麦	2005—2008 年	林科验 2008 年 154 号	84.1	48.0	7	0	2.014
37	2003-4-27	人工林木材的 NIR(近红外)材性预测及增值利用技术引进	中国林科院木材工业研究所	木材加工	江泽慧；许忠允	美国	2003—2006 年	林科验 2008 年 113 号	89.0	430.0	51	15	1.873
115	2005-4-69	竹醋液深度开发关键技术引进	浙江林学院	林产化工	方伟	日本	2005—2008 年	林科验 2008 年 004 号	82.0	90.0	4	2	1.777
23	2002-41	落叶松工程木制材料制造与应用技术引进	北京林业大学	木材加工	申世杰	日本；美国	2002—2005 年	林科验 2007 年 051 号	82.9	65.0	9	4	1.774
		……						……					
103	2005-4-40	林木遗传资源监测、评价与管理技术引进	国家林业局科技发展中心	资源利用	王琦	美国；哥伦比亚	2005—2008 年	林科验 2008 年 018 号	87.3	40.0	2	0	0.296
77	2004-4-59	植物品种图像识别技术引进	国家林业局科技发展中心	林业机械	周建仁	德国	2004—2007 年	林科验 2008 年 124 号	85.8	50.0	3	0	0.244
33	2002-57	树木性状基因芯片分析技术引进	中国林科院林业研究所	林业机械	卢孟柱	美国	2002—2005 年	林科验 2007 年 029 号	84.3	80.0	0	0	0.193
84	2004-C04	单宁酸精加工产品生产工艺创新与示范	中国林科院林产化学工业研究所	林产化工	张宗和	日本	2004—2007 年	林科验 2008 年 130 号	89.9	105.0	0	0	0.186
39	2003-4-29	电磁波式高精度木材含水率检测系统引进	国家林业局哈尔滨林业机械研究所	林业机械	刘少刚	美国	2003—2008 年	林科验 2008 年 152 号	76.1	60.0	0	0	0.160

* 注：项目信息以 DEA 超效率值降序排列

表 4 项目评审中的 I 类与 II 类误差的界定

Table 4 Definition of Type I and Type II errors in project evaluation

项目	项目同行评审打分	
	高于中位数	低于中位数
项目科研效率高于中位数	正确	II类误差（Type Ii Errors）：同行评审打分低于平均水平的项目其科研效率反而高于平均水平
项目科研效率低于中位数	I类误差（Type I Errors）：同行评审打分高于平均水平的项目其科研效率反而低于平均水平	正确

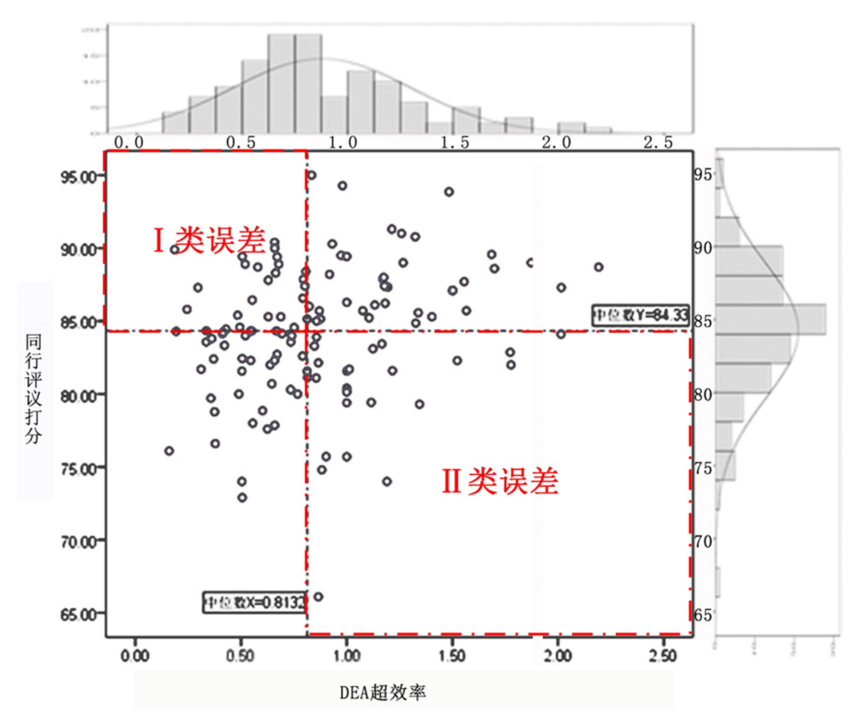


图 3 项目的验收评分与 DEA 超效率值的散点分布

Fig.3 Scattered distribution of scores and DEA super-efficiency value in projects

表 5 项目同行评审犯 I 类与 II 类错误的比重

Table 5 Proportion of Type I and Type II errors in the peer review of projects

项目	数值	比重/%
正确决策	81	64.3
I 类误差	22	17.5
II 类误差	23	18.2
总数	126	100.0

表 6 项目的验收评分与 DEA 超效率（SE）频次的正态分布检验

Table 6 Normal distribution test of the score and DEA super-efficiency (SE) frequency in projects

统计量	Kolmogorov-Smirnov 检验			Shapiro-Wilk 检验		
	统计值 *Statistic*	自由度 *df*	显著性 *Sig.*	统计值 *Statistic*	自由度 *df*	显著性 *Sig.*
DEA-SE	0.102	126	0.003	0.953	126	0.000
Scores	0.070	126	0.200**	0.774	126	0.067**

* 注：**.表示在置信度（双测）为 0.01 时，相关性是显著的

表 7 项目的验收评分与项目典型投入 / 产出指标 Spearman 相关性检验

Table 7 Spearman correlation test between the score and typical input/output indicators in projects

项目	指标	同行评议得分	经费{I1}	期刊论文{O2}	专利{O4}
同行评议得分	相关系数ρ	1.000	0.105	0.250**	0.377**
	显著性*Sig.*	0.000	0.242	0.005	0.000
	样本量*N*	126	126	126	126
经费{I1}	相关系数ρ		1.000	0.129	0.073
	显著性*Sig.*		0.000	0.151	0.416
	样本量*N*		126	126	126
期刊论文{O2}	相关系数ρ			1.000	0.277**
	显著性*Sig.*			0.000	0.002
	样本量*N*			126	126
专利{O4}	相关系数ρ				1.000
	显著性*Sig.*				0.000
	样本量*N*				126

* 注：**.在置信度（双测）为 0.01 时，相关性是显著的

分布并不是正态分布，所以相关性分析，我们不能采用 Pearson 相关性检验而应采用 Spearman 相关性检验。

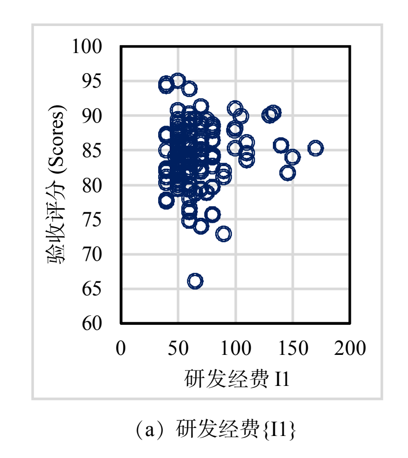
4.3.1 同行评议打分与项目投入 / 产出指标相关性

根据散点分布（图 4）判断与相关性检验（表 7）结果发现：948 林产技术项目验收评分（Scores）与研发经费投入 I1 并不存在明显的相关性，但与期刊论文数量 O2、专利数量 O4 之间存在显著的统计正相关性，Spearman 相关系数分别为 ρ=0.250（*Sig.*<0.01）与 ρ=0.377（*Sig.*<0.01）。这说明：产出越多，同行评分也越高。

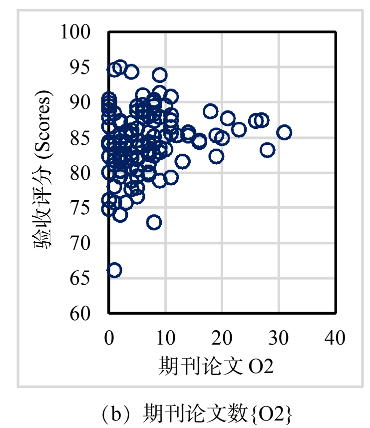
同时，科研产出期刊论文数量 O2 与专利数量 O4 之间也呈现适度的统计正相关性，Spearman 相关系数 ρ=0.277（*Sig.*<0.01）。这与之前一系列研究结论相一致：科研活动学术论文的产出强度与技术专利的产出强度之间存在一致性。

4.3.2 超效率(SE)-DEA 模型指标与项目投入 / 产出指标相关性

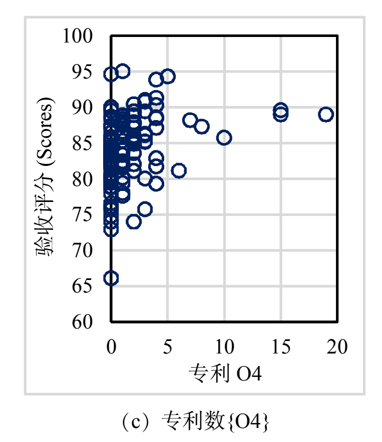
根据散点分布（图 5）判断与相关性检验（表 8）结果发现：948 林产项目的研发效率（DEA 超效率表征）与研发经费 I1 投入之间并不存在明显的相关关系，但它与期刊论文数量 O2、专利数量 O4 之间存在适度的统计正相关性，Spearman 相关系数 ρ=0.376



（a）研发经费{I1}



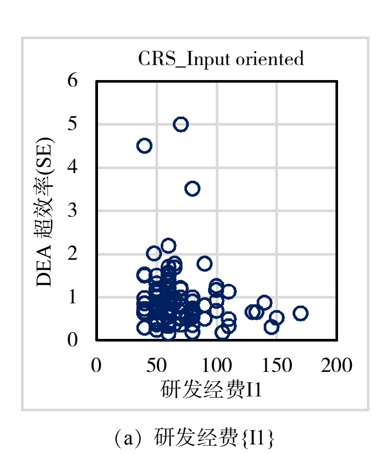
（b）期刊论文数{O2}



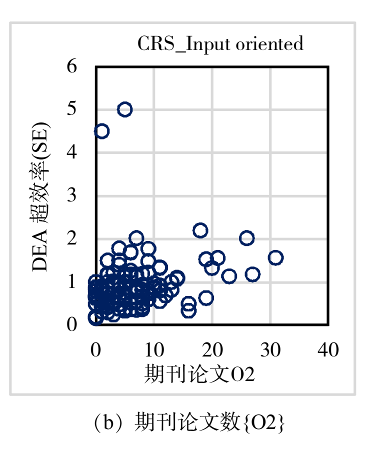
（c）专利数{O4}

图 4 项目的验收评分与项目典型投入 / 产出指标的散点分布

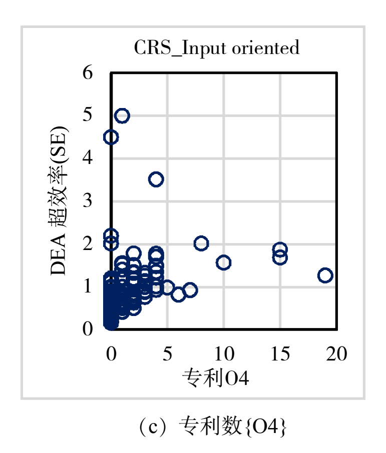
Fig.4 Scatter distribution of the score and typical input/output indicators in projects



(a) 研发经费{I1}



(b) 期刊论文数{O2}



(c) 专利数{O4}

* 注：本文的 DEA 超效率（SE）采用的是规模报酬不变（CRS）的 CCR 模型并以投入角度（Input Oriented）进行计算

图 5 项目 DEA 超效率（SE）与项目典型投入 / 产出指标的散点分布

Fig.5 Scatter distribution of the DEA super efficiency (SE) and typical input/output indicators in projects

表 8 项目的 DEA 超效率（SE）与项目典型投入 / 产出指标 Spearman 相关性检验

Table 8 Spearman correlation test of the DEA super efficiency (SE) and typical input/output indicators in projects

项目		DEA Super Efficiency	经费{I1}	期刊论文{O2}	专利{O4}
DEA Super Efficiency	相关系数ρ	1.000	-0.099	0.376**	0.523**
	显著性*Sig.*	0.000	0.271	0.000	0.000
	样本量*N*	126	126	126	126

* 注：**.在置信度（双测）为 0.01 时，相关性是显著的

(*Sig.*<0.01) 与 ρ=0.523 (*Sig.*<0.01)。这说明：产出越多，项目生产效率也越高。

3.3.3 同行评议结果与超效率(SE)-DEA 模型指标相关性

根据散点分布（图 6）与相关性检验（表 9）结果发现：林业 948 产业技术项目验收评分与研发（R&D）主体的科研效率（DEA 超效率表征）之间呈现统计正相关性，Spearman 相关系数 ρ=0.250（*Sig.*<0.01）。这说明：同行评议得分与项目的研发效率测度结果之间存在一致性。

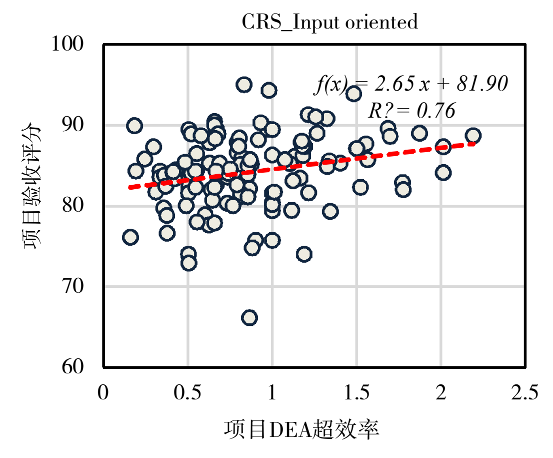


图 6 项目的验收评分与 DEA 超效率的散点分布

Fig.6 Scattered distribution of the score and DEA super-efficiency in projects

* 注：本文超效率以投入角度采用规模报酬不变（CRS）的 CCR 模型进行计算；曲线的拟合采用最小二乘（OLS）方法计算，红色虚线为拟合曲线

5 结论与讨论

5.1 结 论

本文借助于 DEA 方法评价决策单元（或 DMU）的研发生产率（效率），并将该科研生产率（效率）指标作为同行评议决策有效性的验证标准，以求通过后验性地核查方式了解整个同行评议决策的质量。实验数据选择国家林业局在“十五”期间批准实施的 126

表 9　项目的验收评分与 DEA 超效率的 Spearman 相关性检验

Table 9 Spearman correlation test between the score and DEA super-efficiency in projects

项目		DEA_Super Efficiency	Scores
DEA_Super Efficiency	相关系数ρ	1.000	0.250**
	显著性*Sig.*	0.000	0.005
	样本量*N*	126	126

* 注：**.在置信度（双测）为 0.01 时，相关性是显著的

项“948 计划”林业产业技术项目，实验模型为超效率 SE-CCR 模型，研发投入 9 项，研发产出 4 项，测度研发效率。

与以往以引用为基础的量度指标作为同行评议决策有效性的验证标准相比较，本文的实验研究结果得出的结论相似，即专家组的同行评议质量判断与实际研发（R&D）效率（DEA 超效率）二者之间呈现适度的统计正相关性，Spearman 相关系数 ρ=0.250（*Sig.*<0.01）。同时，犯 I 类错误（高估）与 II 类错误（低估）的比例分别为 17.5%与 18.2%，有效同行评议决策比重仅占 64.3%。该结果证实：同行专家组的学术质量价值判断的确存在系统误差但规模有限，实际的项目资金分配并没有我们想象地那样高效。该比例与西欧发达经济体项目资助委员会大型基础研究项目资助评审的有效性验证结果有一定差异，主要在于西欧的项目资助评审更倾向于犯 II 类错误（弃真）而非 I 类错误（纳伪），即为实现高水平项目的有效遴选“宁可错杀一千也不放过一个”，欧美的项目资助评审更为严格也更为完备、系统。这对国内基金委的项目运作的建议：加强项目“过程质量监管”，严格层层把关，扭转中国当前“重立项轻验收”的局面，就如习总书记指出的由“高速发展”向“高质量发展”转型。

5.2 讨　论

我们并不认为该评价结论普遍适用且一成不变，但是本文的确以准确数字来阐释同行评议质量到底好到“何种程度”。未来研究可能会以更全局、更大规模、更长周期的统计样本数据源开展实证分析，论证同行评议的质量在长周期是否不断完善？或者通过多种客观量度方法对同行评议的有效性获得科学性地深入理解，又或者进一步更为直观地对比多种量度工具的具体价值。

参考文献：

[1] PRICE D J D. Little science, big science–and beyond[M]. New York: Columbia university press, 1986.

[2] 万昊, 谭宗颖, 朱相丽. 同行评议与文献计量在科研评价中的作用分析比较[J]. 图书情报工作, 2017, 61(1): 134–152.

WAN H, TAN Z Y, ZHU X L. Comparison of the role of peer review and bibliometrics in research evaluation[J]. Library and information service, 2017, 61(1): 134–152.

[3] MERTON R K. Social theory and social structure [M]. New York: Free press, 1968.

[4] 姜春林, 张立伟, 刘学. 中外同行评议研究现状及问题探讨[J]. 科技管理研究, 2015(3): 163–166.

JIANG C L, ZHANG L W, LIU X. The status quo of peer review research and its problems[J]. Science and technology management research, 2015(3): 163–166.

[5] 盛怡瑾, 初景利. 同行评议质量控制方法研究进展[J]. 出版科学, 2018, 26(5): 48–55.

SHENG Y J, CHU J L. Research progress of peer review quality control method[J]. Publishing journal, 2018, 26(5): 48–55.

[6] CARTER B. Peer review a good but flawed system[J]. Journal of child health care, 2017, 21(3): 233–235.

[7] WESSELY S, WOOD F. Peer review of grant applications: A systematic review[M]//, GODLEE F, JEFFERSON T. Peer review in health sciences, London: British medical journal books, 1999.

[8] MAVIS B, KATZ M. Evaluation of a program supporting scholarly productivity for new investigators[J]. Academic medicine, 2003, 78 (7): 757–765.

[9] BORNMANN L, DANIEL H D. Convergent validation of peer review decisions using the h index. Extent of and reasons for type I and type II errors[J]. Journal of informetrics, 2007, 1(3): 204-213.

[10] BORNMANN L, WALLON G, LEDIN A. Does the committee peer review select the best applicants for funding? An investigation of the selection process for two european molecular biology organization programmes[J]. PLoS ONE, 2008, 3(10): E3480.

[11] NEUFELD J, VON INS M. Informed peer review and uninformed bibliometrics?[J]. Research evaluation, 2011, 20(1): 31-46.

[12] LI D, AGHA L. Big names or big ideas: Do peer review panels select the best science proposals?[J]. Science, 2015, 348(6233): 434-438.

[13] MELIN G, DANELL R. The top eight percent: Development of approved and rejected applicants for a prestigious grant in Sweden[J]. Science and public policy, 2006, 33(10): 702-712.

[14] HORNBOSTEL S, BÖHMER S, KLINGSPORN B, et al. Funding of young scientist and scientific excellence[J]. Scientometrics, 2009, 79(1): 171-190.

[15] VAN DEN BESSELAAR P, LEYDESDORFF L. Past performance, peer review, and project selection: A case study in the social and behavioral sciences[J]. Research evaluation, 2009, 18(4): 273-288.

[16] BORNMANN L, LEYDESDORFF L, VAN DEN BESSELAAR P. A meta-evaluation of scientific research proposals: Different ways of comparing rejected to awarded applications[J]. Journal of informetrics, 2010, 4(3): 211-220.

[17] BUXTON M, HANNEY S, MORRIS S, et al. Medical research - What's it worth? Estimating the economic benefits from medical research in the UK[R]. Health economics research group, office of health economics & rand, Europe, London: UK evaluation forum, 2008.

[18] MACILWAIN C. Science economics: What science is really worth?[J]. Nature, 2010, 465(10): 682-684.

[19] 王海宁, 李姗姗, 栾贞增. 高校科研成果转化能力与效率关系评价——基于 2016 年教育部直属 61 所高校的实证研究[J]. 科技管理研究, 2018, 406, 38(12): 147-154.

WANG H N, LI S S, LUAN Z Z. Evaluation on the relationship between transformation ability and efficiency of scientific research achievements in colleges and universities: An empirical study of 61 universities directly under the chinese ministry of education in 2016[J]. Science and technology management research, 2018, 406, 38(12): 147-154.

[20] HAZELKORN E. European commission: Expert group on assessment of university-based research, assessing Europe's university based research[R]. Dublin institute of technology, centre for social and educational research, 2010.

[21] IVANOVA I A, LEYDESDORFF L. Knowledge-generating efficiency in innovation systems the acceleration of technological paradigm changes with increasing complexity[J]. Technological forecasting & social change, 2015, 96: 254-265.

[22] SALTER A J, MARTIN B R. The economic benefits of publicly funded basic research: a critical review[J]. Research policy, 2001, 30(3): 509-532.

The Validity of Peer Review Results of DEA Based Super Efficiency Projects

WAN Hao[1], ZHANG Fujun[2], LV Qianqian[3]

(1. Library of Shandong University of Science and Technology, Qingdao 266590;

2. Development Planning Office, Shandong University of Science and Technology, Qingdao 266590;

3. Document Information Center, Chinese Academy of Sciences, Beijing 100190)

Abstract: [Purpose/Significance] This article uses some projects'peer review results as an example to explore the effectiveness of peer review based on expert knowledge. [Method/Process] It uses data envelopment analysis to measure the R&D efficiency of the projects. The evaluation model is a CCR model modified by super-efficiency, and the efficiency index is used as a verification standard for the effectiveness of peer review results. The data source is 126 "948 projects" implemented by the State Forestry Administration, covering 13 input and output evaluation indicators. [Results/Conclusions] The experiment found that there was a moderate positive correlation between the project peer review results and the actual R&D efficiency (DEA super efficiency), and the Spearman correlation coefficient ρ=0.250 (*Sig.*<0.01). The percentages of type I errors and type II errors made by subjective review were 17.5% and 18.2%, respectively, and effective peer review decisions accounted for 64.3%. It is confirmed that there are certain systematic errors in peer review, but they are inevitable.

Keywords: peer review; project evaluation; effectiveness; super efficiency DEA